INTRODUCTION TO DATA-DRIVEN RELIABILITY ENGINEERING

VALERI VOEV

WHAT IS RELIABILITY

- Reliability is defined as the probability that a device (system) will perform its *intended function* during a *specified period of time* under stated conditions
- Let us denote by T the lifetime of the system under consideration. As we live in a world where design, manufacturing, transport and operation cannot be held perfectly constant (or perfectly controlled), T will be a *random* quantity
- We can now define reliability more formally as $R(t) = \Pr[T > t]$, that is the reliability of a system at time t is the probability that the system's lifetime will exceed t (or as stated above will perform its intended function at least as long as t). $1 - R(t)$ is then the probability that the system will **not** perform its function and we typically call this the failure rate $F(t)$.
- (Data-driven) reliability engineering is about studying, estimating and analysing reliability of components and systems (using lifetime and other data).

WHY DO WE NEED RELIABILITY ENGINEERING

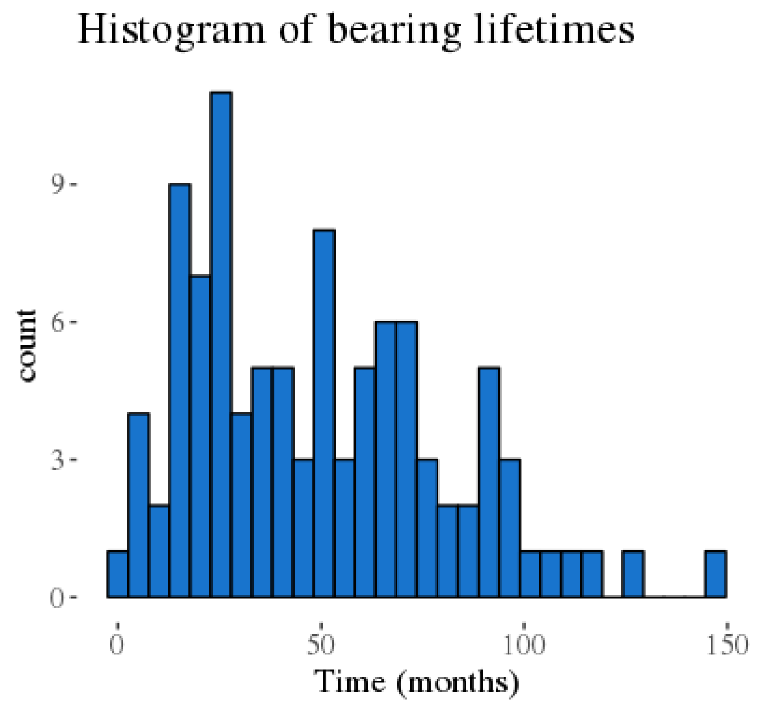
- As operators or owners of assets, reliability is a key input to lifecycle costs
- As designers and engineers reliability is a key design parameter which will determine the cost of the component
- Knowing the reliability of a system is a key prerequisite to design improvements
- Setting up test plans requires an understanding of reliability theory to make sense of the results – what reliability did a test prove, for how long (or how many units) I need to test to be able to prove a given reliability level?
- Reliability can (and should) be used to guide maintenance decisions – if unexpected (unscheduled) replacements are costly, is there an optimal replacement period (and what is it) which optimizes overall cost? The answer to this question will depend on the reliability and replacement cost of the unit.
- Reliability is at the heart of Reliability Centered Maintenance

LET’S DIG IN…

- Assume we test 100 bearings to failure and record the time to failure on each test. The first 8 failure times are (let’s say we are measuring in equivalent months from an accelerated lifetime test):

Time to failure (months)
18.81
52.09
63.43
63.74
16.29
59.30
27.69
41.96

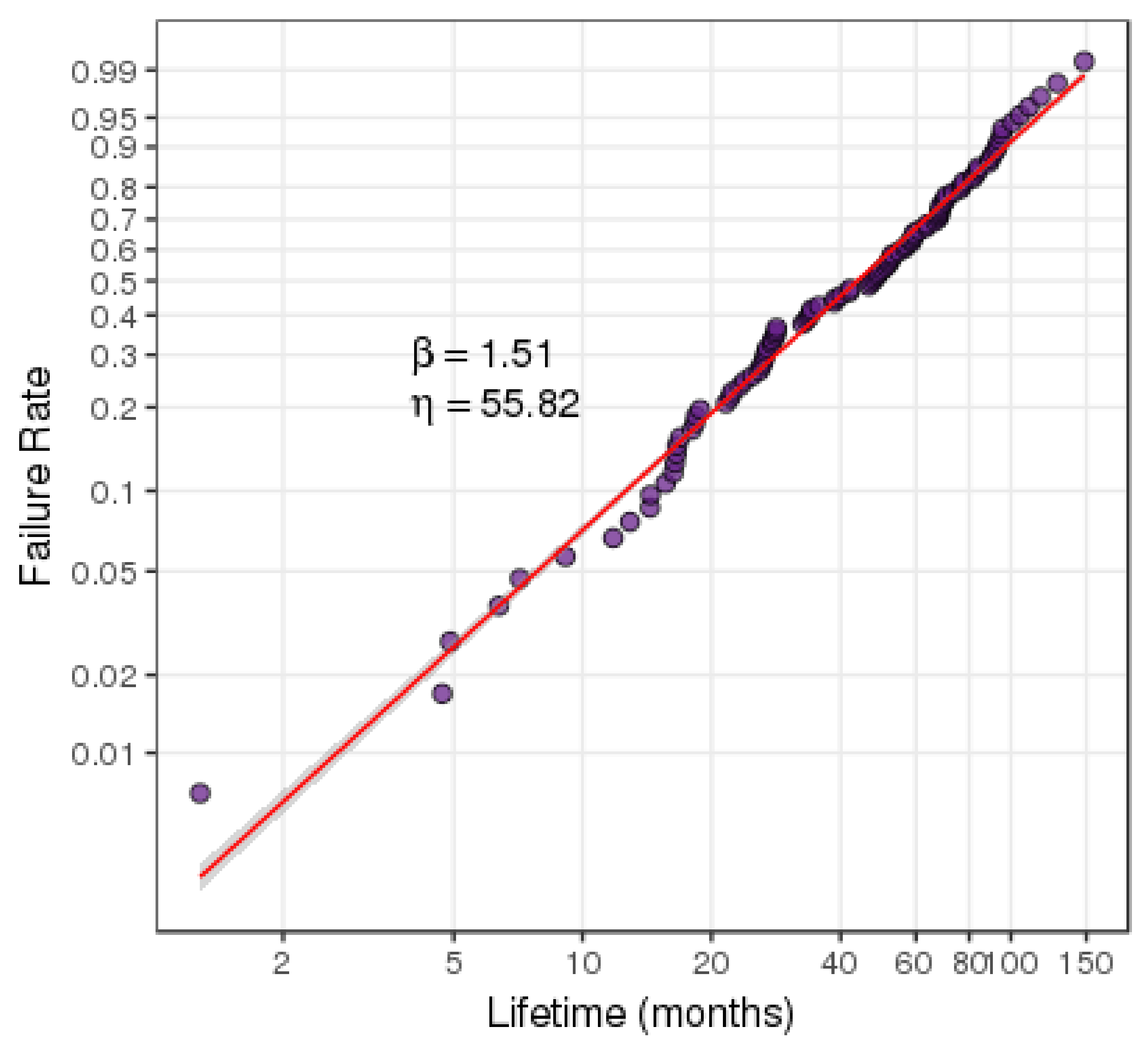
- The histogram of all 100 values is given below



- What is the reliability of this bearing at 5 years (60 months) and at 10 years (120 months)?
- What is the L10 life (the time at which the failure rate is 10%)
- What is the 90% confidence bound for the calculations above?
- If this bearing is designed for a lifetime (L10) of 2 years did this test confirm or reject this design?

WEIBULL ANALYSIS

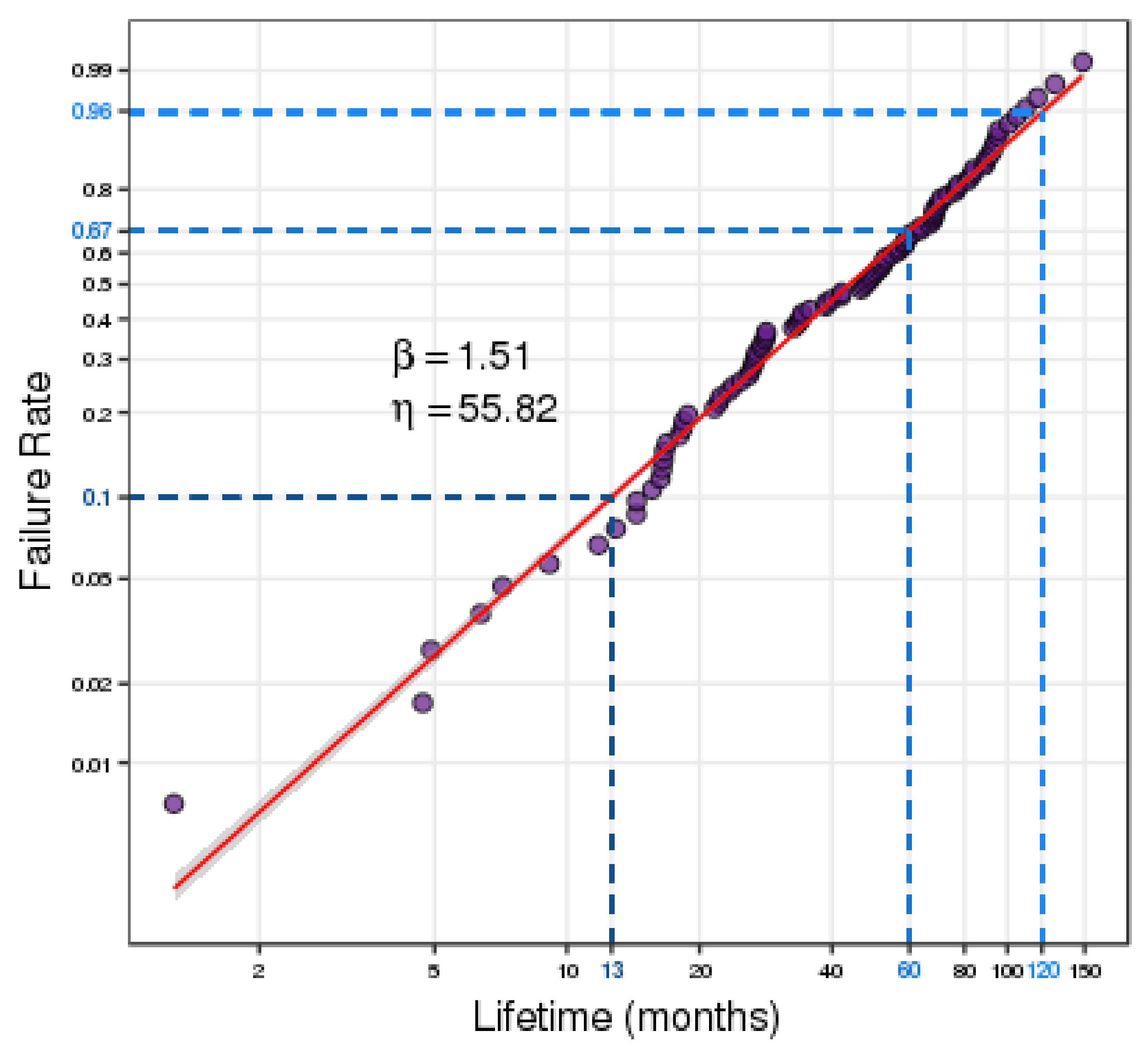
- The most common statistical technique for analysing lifetime data is Weibull analysis due to it's flexibility and ease of interpretation (few parameters)
- The Weibull distribution has 2 parameters, shape (β) and scale(η) and its distribution function is given by $F(t) = 1 - \exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right)$
- The parameters η and β have important interpretations: η is also called characteristic life and determines the time by which 63.2% of all units will have failed, while β also called slope parameter is indicative of the failure mode:
  - β < 1 indicates infant mortality (e.g, in electronic components due to poor quality control, mis-assembly, etc)
  - β = 1 indicates random failure independent of component age (e.g. human errors, natural causes, etc)
  - β > 1 indicates wear-out type of failure (failure rate increases with age, e.g. fatigue, corrosion, erosion)
- Once we have a way to estimate β and η we can calculate component reliability at any time t



- This is a so-called Weibull plot
- Using log transformations, the Weibull CDF can be linearized and the β and η can be estimated using a linear regression – this method is called median rank regression
- There are other methods (e.g. maximum likelihood) to estimate these parameters

This plot tells us some important things:

- About 2/3 of all bearings will fail by 56 months of operation (that follows directly from the value of η)
- The failure mode is a slow wear-out likely due to low-cycle fatigue (that follows from the value of β)



- The L10 life is 13 months which means that the bearing doesn't conform to the design specifications
- 5-year reliability is ≈33% and 10-year reliability is ≈4%
- Confidence bounds (grey area around the linear fit) are very tight since we have a decent amount of data

Implications for maintenance strategy:

- Consider including a replacement of these bearings at an annual frequency (depending on the cost of unscheduled repair)
- From a cost-benefit analysis is there (a potentially more costly) bearing design with better reliability which will decrease overall CAPEX + OPEX costs?

EXTENSIONS

- In many cases your data will not look so nice on Weibull paper. That suggests that you might need to consider:
  - Different distribution – e.g. Lognormal
  - 3-parameter Weibull (e.g., in cases where failure-free time is a reasonable conjecture)
  - Multiple failure modes (bathtub curve)
- In our example we worked with time as the measure of lifetime. Different measures such as cycles, miles, production (e.g. for wind turbines) can be used in a completely analogous way
- With field data, you will have “suspensions” or units which have not failed at the time of data collection, also called (right-) censored data – the Weibull methodology can easily handle that
- Often, various units can be exposed to very different environments (e.g., wind turbines are placed in very different conditions with respect to turbulence, shear, ambient temperature – we mention a class of models that can handle this on the next slide

PROPORTIONAL HAZARD MODELS

- The concept of *hazard rate,* $\mathrm{h(t)}$, is central in survival analysis. The hazard rate can loosely be defined as the probability of failure over the next small period of time $[t; t + \Delta t]$ given that the system has not failed up to time t. For example, for the Weibull distribution, the hazard rate takes the form

$$\mathrm{h(t)} = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$$

- In proportional hazard models the hazard rate is modelled as a function of some exogenous variables, $X_1, X_2, \dots, X_p$ as $\mathrm{h(t)}$ = $\mathrm{h_0(t)exp}(\lambda_1 X_{1t} + \lambda_2 X_{2t} + \cdots + \lambda_p X_{pt})$. $\mathrm{h_0(t)}$ is called the baseline hazard (could be the Weibull hazard given above).
- The λ parameters (together with the parameters in the baseline hazard) are estimated using maximum likelihood. For example, if X_{1t} is the ambient temperature at time t, and λ_1 is positive, then a higher ambient temperature increases the hazard rate (and failure rate).

If units are exposed to different ambient temperature in the field, the proportional hazard model will enable us to reflect that in their reliability analysis and consequently lead to customized maintenance strategy!

SOFTWARE

- A number of commercially available packages can estimate and provide further functionality based on Weibull analysis. Some are Reliasoft, Super Smith, as well as industry specific packages (e.g. DNV GL's Maros for the oil and gas industry)
- The examples in these slides are made from scratch in R (which is open source) – so one doesn't need a big investment in software
- Much more important than software is data collection, organization, storage and access. In many cases, data is collected in a decentralized, non-standardized way which presents significant challenges to a data-driven reliability-oriented culture